The Importance of Identifying the Right Data Collection to Facilitate a Better Decision Making

Che Rosnah Ngah
Universiti Malaysia Terengganu
cherosnah@umt.edu.my

Abstract

In a knowledge management setting, data is placed at the bottom of the hierarchy as the initializer before data is converted to the next hierarchy level which is called information. Although data and information are easily available, their usefulness may diminish if they are not handled carefully. This may be due to the method of data collection or the degree of its originality. A researcher is aware that the decision making process is based on having both facts and figures which are complementary to each other. However sometimes, people have a tendency to keep lots of unused data for the sake of 'just-in-case' needs. This happens when precise needs are not properly identified and creates wastage of energy and resources, not to mention improper decision making. In the knowledge based society that we are aiming for, data and information must be used optimally and a good knowledge repository must be created. A solution to acquire a good set of storage and retrieval data, apart from prior determination, is to have proper skills of data interpretation and analysis using appropriate ICT tools which include statistical sampling method. The ultimate aim is to overcome information explosion where people are drowning in information but starving for knowledge. Examples are outlined showing how data can be manipulated to reflect false pictures due to wrong decision making based on meaningless and doubtful non-quality data. Knowledge workers are responsible for embedding prior value the moment data is collected or used. Proper actions taken by the knowledge workers will result in a good knowledge repository which contributes positively towards a knowledge society.

Abstrak

Dalam persekitaran pengurusan pengetahuan, data ditempatkan di aras bawah dalam hirarki sebagai permulaan sebelum ia diubah ke hirarki seterusnya yang dinamakan maklumat. Walaupun data dan maklumat kadangkala mudah didapati, kegunaannya akan berkurangan sekiranya tidak diuruskan dengan betul. Ia mungkin kerana kaedah pencarian data atau tahap keaslian data itu sendiri. Penyelidik menyedari bahawa proses membuat keputusan masakini dibuat berdasarkan fakta dan perangkaan yang saling melengkapi antara satu sama lain. Walau bagaimanapun, kadangkala seseorang itu cenderung untuk menyimpan data yang tidak digunakan. Hal ini terjadi kerana keperluan yang tepat tidak dikenalpasti yang menyebabkan pembaziran tenaga dan sumber. Ini tidak termasuk kesan daripada membuat keputusan yang tidak tepat. Di dalam masyarakat berasaskan pengetahuan yang kita ingin capai, data dan maklumat mestilah digunakan secara optima dan repositori ilmu pengetahuan yang baik perlu dihasilkan. Penyelesaian untuk mendapatkan set storan dan capaian data ialah untuk mendapatkan kemahiran interpretasi data dan analisis menggunakan peralatan ICT yang mengandungi kaedah persampelan secara statistik. Sasaran yang penting adalah untuk mengatasi era ledakan maklumat di mana seseorang itu mencari maklumat dengan mendalam tetapi ketandusan ilmu pengetahuan. Contoh-contoh yang diberikan adalah garis panduan yang menunjukkan bagaimana data boleh dimanipulasi dan berkait dengan gambaran yang salah kerana membuat keputusan yang salah berdasarkan data yang tidak mempunyai makna dan tidak berkualiti atau menimbulkan keraguan. Pekerja yang terlibat di dalam ilmu pengetahuan bertanggungjawab memasukkan nilai yang asal semasa data dikumpul atau digunakan. Tindakan yang sepatutnya diambil oleh mereka ini akan menghasilkan repositori ilmu pengetahuan yang dapat menyumbang kepada masyarakat berilmu pengetahuan secara positif.

Introduction

Why is data important? Does data give any information? To answer these two questions, understanding what data is all about and the difference between data and information is required. In some cases, data and information are used interchangeably although there is a great difference in meaning between the two words. In knowledge management, data is placed at the bottom of the hierarchy as the main initial starter before data is converted to the next level which is known as information. Identification in the early stage of data existence is vital to indicate as to what degree it can be used. Lack of proper identification of data usage will result in wastage of resources, energy and time taken in generating data to information.

Data is usually collected for either anticipation or curiosity reasons which unconsciously deals with the decision making process. Data might be collected as much as it can be gathered for 'just-in-case' needs. This may result in lots of redundant data which may not be used at all. Therefore, it is crucial to start with the correct steps.

A sound organization is usually made up of good knowledge workers where they are able to determine from the beginning how data can be collected in order to facilitate better decision making. The person, therefore, needs to possess skills of understanding the purpose of collecting data and must be able to foresee the end requirements. Otherwise, appropriate matching of wish-to-be results may not be attained.

When data is wrongly collected, the interpretation made may mislead and defeat the purpose. In the end, results that are being reported may give zero contribution to the strategic planning of the whole organization. Thus, the community it serves indirectly experience a benefit loss. Imagine in an academic setting where the university extends its functions to the community becomes distorted due to wrong decision making. There will be a lot of wastage in terms of human resources capabilities, physical infrastructure, time and effort. The best synergy is to have incorporated simple data collection method with clear purpose of collecting as this will help in tailored decision making. During the process, ICT applications can be used to acquire error-free valid calculations and easy process expedition.

This paper highlights an example taken from an academic library's service, that is, data collected for the library information literacy class offered by the library. The example also demonstrates that data can sometimes be manipulated even when a quality objective has been set. In this case, librarians who are also knowledge workers are responsible in generating good data collection before it can be converted to give its meaning to become information. Thus, information will provide knowledge which would contribute to supporting the learning, teaching and research of the academic community.

Data in Action

The definition of data is "things known, or assumed facts and figures, from which conclusions can be inferred" (Taylor & Cihon, 2004). Data is also said to be systematically recorded information, whether numbers or labels, together with its context (De Veaux; Velleman & Bock, 2005). Griffiths (1998) even mentioned that data is more than numbers whereby numbers themselves can hold information that is important.

Understanding Data via Knowledge Hierarchy

Figure 1 highlights the position of data, information, knowledge and intelligence with its value intensity in a pyramid format. Each level is embedded with its justification before each hierarchy travels from low to high value. Data which is assembled with facts, observations and data points will be of higher value as information after context is added. Information with context will add further value to become knowledge. When knowledge is paired with meaning, the value increases to become intelligence or wisdom where an insight has been embedded.

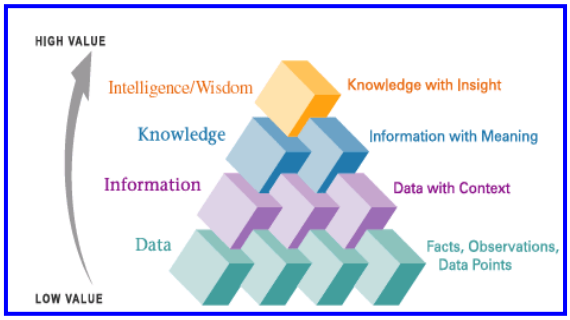

Figure 1: Knowledge Hierarchy
(Source: www.transknowformance.com)

Knowledge workers are the people who will be responsible for handling the flow of data from its low to high value.

Why Knowledge Workers?

They are named knowledge workers as they work closely with knowledge activities apart from being professionals in the field. Researchers, librarians, academicians or data interpreters/keepers are among the people who belong to this category. Peter Drucker, a famous management guru, had coined the term in his book entitled *Landmarks of Tomorrow* way back in 1959, where he had foreseen how knowledge workers can be a great human asset in an organization. Without their role and expertise, it could jeopardize any intelligent business decision in an organization as data will be initially handled by them.

Data with Context Value

De Veaux (2005) stated that data values are useless without their context. In order for good knowledge workers to spell out clearly data with context, the "Five W's: Who, What, When, Where, Why and How" can be established. Answering these questions can provide context for data values especially the Who and What which are essential. Failing to answer these two elements will result with no data obtained and thus no useful information can be harvested.

Data production itself must be planned (Taylor & Cihon, 2004). The complexity of modern measurement requires a considerable amount of planning to ensure data are meaningful. It has been highlighted that any gaps in a database resulting from omissions or data rejection can weaken the conclusions and even make decisions impossible in certain cases.

Let us look at an example in which a library receives a request from lecturers to conduct information literacy classes as outlined in Table 1. An information literacy class refers to a service offered by the library to share skills in information searching techniques. The class is usually tailored to a request made and is conducted in a small recommended size of not more than 30 people. Classes run in the computer labs where students will do a live search together with a presenter (librarian) on a specific theme or subject agreed upon. A few facilitators (librarians) may assist to expedite students' understanding and at the same time, their role is to ensure data collection on class progress is gathered through the distribution of student evaluation forms at the end of the class.

Table 1: Data of library information literacy class request

Requester (s)	Lecturer A	Lecturer B	Lecturer C	Lecturer D
Students category	Year 1	Year 2	Year 4	Year 4
Faculty	Mixed	Business	Science	Business
Subject covered	Public speaking	Management	Engineering	Accounting
No. of students	200	120	46	75
Time slots required	8	4	2	3
Duration of class	1 hour	1 hour	1 hour	1 hour

Lecturers A, B, C and D **who** made the request had justified **what** sort of theme is required for **how** many number of students under their care. In return, librarians identify **who** are the audience; **when and where** to be conducted**, why** it is needed and **what and how** to measure service performance.

Based on these data, librarians list the following process actions:

- Identify how many time slots to be held and tabulate schedule.
- Determine how many assistants needed for each session and who they are .
- Plan for appropriate venue, technical help and setting requirements.
- Think how delivery can be measured in terms of students' satisfaction rate.
- Match end results to be congruent to the objective quality set.
- Prepare the necessary aids to assist students' understanding.
- Prepare for the day.

It may look simple but a focused and know-what-to-do knowledge worker should be able to think quickly in the sequence shown in figure 2:

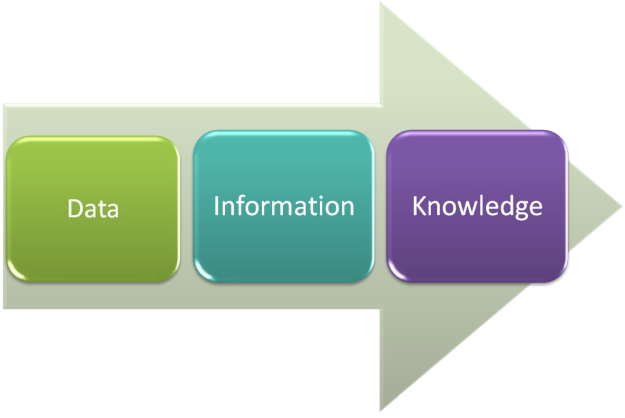

Figure 2: Knowledge sequence

This is because in an academic library environment, it is very important that any activity concerning teaching and learning must be complimentary to the whole system of higher learning. Librarians need to think that service given to students and lecturers is part of their contribution towards the library's community.

Converting Data to Information

In this case, librarians may incorporate a student evaluation form which serves as a data collection tool for students' satisfaction. Again, questions that students need to respond are constructed to provide quantifiable answers (Table 2).

Demographic data such as name, gender and age are not included as these information will not be analysed. However, data on faculty and year of study could help for future planning. By identifying the end purpose, no unused data is collected. This could save space, time and energy in making decisions even the decision to plan the class itself initially.

Table 2: Library information literacy class measurement

No	Questions	Rating Scale				
1.	Achieve class objective	1	2	3	4	5
2.	Additional information after attending class	1	2	3	4	5
3.	Content of module is appropriate with time allocation	1	2	3	4	5
4.	Hands-on exercise assist students in understanding information searching techniques	1	2	3	4	5
5.	Comfortable venue	1	2	3	4	5
6.	Presenter's delivery easy to understand	1	2	3	4	5
7.	Facilitator(s) is/are willing to help and facilitate learning	1	2	3	4	5

At the end of the information literacy class, facilitator(s) will distribute student evaluation forms to be filled up by students before they leave the class. A Likert measurement with a rating scale of 1 to 5 is tabulated and students need to circle the score that they deemed suitable. The numbers are then translated as follows:

1 - Poor
2 - Satisfactory
3 - Good
4 - Very Good
5 - Excellent

Each question is analyzed individually and necessary actions will be taken to address those unsatisfactory achievements. In addition, attempt is also made to address the open ended questions which are subjectively analyzed as respondents are requested to provide comments or suggestions for improvements in their own words.

Data Used for Investigation

After giving representation to the above scale, they are used for investigation whereby statistics are generated. Agresti and Franklin (2007) stated that statistics is the art and science of learning from data. According to them, when data is used for investigation, data is not only being translated into knowledge (after it has been given a meaning as shown in Figure 1), but it is also where understanding the world is taking place. Results obtained after proper analysis done will tell whether it is successful or not. Decision making will be based on factual/tested data although prior target or hypothesis could have been constructed.

Some hypothesis statements could be:

a) Competent presenter will draw a high percentage of customers' satisfaction;
b) Comfortable venue facilitates conducive learning;
c) Right module contents make effective learning;
d) Hands-on exercise provides real experience in searching techniques.
e) Well-prepared facilitators contribute to learning experience.

The assumptions made therefore are tested quantitatively before the final decision is concluded.

Transforming Information into Knowledge

Difference between Information and Knowledge

Jashapara (2004) quoted from Meadow (2001) that information could be considered as systematically organized data. He also stated that another conception of information is data that is endowed with meaning (not necessarily having a scientific meaning), relevance and purpose. Information gives shape to the data and makes a difference to the outlook or insight of the receiver of the data.

Jashapara (2004) again outlined that knowledge could be considered as actionable information which allows us to make better decisions and provides effective input dialogue and creativity in organizations. Tiwana (2000) quoted by Jashapara (2004), also said that this occurs by providing information at the right place, right time and in appropriate format, which is what knowledge management requirement is. Knowledge which is more complex allows us to act more effectively than information or data and equips us with a greater ability to predict future outcomes.

Quality Objective versus Data Manipulation

The quality objective set for each library information literacy class is to achieve at least 80% customer satisfaction. However, the 80% target set may not necessarily reflect the effectiveness of the class delivery. Data manipulation may occur due to several reasons such as:

a) Respondents' biasness/favoritism-students choose presenter/facilitator(s) of their choice
b) One time measurement - analysis is based on one particular show which obviously has certain limitations such as internet access, room condition or lack of team preparation.
c) Key-in errors-human typo errors while inputting data in a hurry.
d) Uncertain respondents or indecisiveness - difficulty in giving quick response/could not think what should be responded.

The target customers' satisfaction rate is just sufficient to obtain minimal feedback of how a service is performing. Achieving 80% or more will definitely reflect the effort made in holding the said class is successful. If the percentage is less than 80% of the customers' satisfaction rate, it is an alarming indicator that something should be done to close the gap. In a sound quality management system such as ISO 9001:2000, TQM or Kaizen, any discrepancies/unmet objectives will need to be addressed right from the root cause.

ICT applications can be a good supporting tool. Technologies are a means to link functional areas providing effective conduit for information and the sharing of knowledge. People within the organizations must be well trained in the use and understanding of technology for maximum benefits to be gained (Mitchell, 2003). This is also to avoid data manipulation affecting the objective set which can influence strategic decision making.

Lessons Learned

a) Effective learning - Although there are only seven questions tabulated quantitatively in the simple student evaluation form, data collection has provided a degree of measurement. This should enable librarians to identify students' behavior towards learning. Effective learning is not a matter of the right attitude or motivation (Argyris, 1998) but rather it is a product of the way people reason about their own behavior. This indicates that knowing customers' profile provides a competitive edge for the library marketing services by knowing what should be prepared or provided to match the customer' needs. The final aim is not only to meet such needs but to exceed the customers' requirements which serve as a contribution to the whole community.

b) Contribution to community - Collecting simple data of the library information literacy class shows that the decision making process is sometimes not easy. It has been accepted that assumptions can be used at times but the accuracy is doubtful. Bias may slip in without notice, thus affecting the contribution to the community it serves. The basic assumption is not always true without data supporting it. Even when figures are used for results analysis, it could not be continually used for future implications (dePaula & Fischer, 2006). Learning from the past is always good but the same answers are not to be expected. Knowledge of the past has been said by George Santayana as "those who cannot remember the past are condemned to repeat it".

The community, who are the university students, will benefit from good data analysis. For example, if the speaker's delivery is not up to the expectation, the library will have to ensure that the staff concerned will will be taught presentation skills. Or if the module content is inappropriately constructed, a review should then be done to revamp the flaw. The final objective is not only to maintain but to also achieve more than what is targeted.

Conclusion

In an academic organization, good knowledge management practice must be created in line with its teaching, learning and research processes. Many knowledge workers are available in different fields to expedite the process of best practice of knowledge management (Dani et.al., 2006). This is crucial as capturing and reusing knowledge has been identified as one of the requirements for the next-generation product development. The library serves as a supporting service in the university and plays a vital role in helping this aim to be achieved. An example of collecting data for further analysis of a library information literacy class is seen to be a minimal consideration of how useful the right data have to be collected. Interpretation and analysis which is embedded in the decision making process may yield expected answers or strictly unpredictable surprising results. This will depend on how to ask the right questions in order to receive the right answer. Some data filtration need to be considered as to avoid wastage of time, energy and resources.

Therefore, information could be saved and deposited into the knowledge repository/base for future fair use using ICT support. Thus, sharing and retrieval of knowledge can be exchanged effectively within the knowledge vicinity.

References

Agresti, A., & Franklin, C. (2007). *Statistics: The art and science of learning from data.* New Jersey : Pearson Prentice Hall.

Argyris, C. (1991). Teaching smart people how to learn. *Harvard Business Review , 69*(3), 99-109.

Dani, S. et al. (2006). A methodology for best practice knowledge management. *Proceedings of the Institution of Mechanical Engineers-Part B-Engineering Manufacture,* 220(10), 1717-1728.

De Veaux, R.D., Velleman, P.F., & Bock, D.E. (2005). *Statistics: Data and Models.* Boston : Addison Wesley.

Griffiths, D., Stirling, W.D., & Weldon, K.L. (1998). *Understanding Data: Principles & Practice of Statistics.* Brisbane : John Wiley & Sons.

Jashapara, A. (2004). *Knowledge management: An integrated approach*. Harlow : Prentice Hall.

Mitchell, H.J. (2003). Technology and knowledge management: Is technology just an enabler or does it also add value?. In E. Coakes (Ed), *Knowledge management: Current issues and challenges* (pp.66-78). Hershey : IGI Publishing.

dePaula, R., & Fischer, G. (2006). Knowledge management: Why learning from the past is not enough!.In J. Daris, E. Subrahmaniam & A. Westerberg (Eds.), *Knowledge management: Organizational and technological dimensions* (pp.21-54). Heidelberg : Springer-Verlag.

Taylor, J, K., & Cihon, C. (2004). *Statistical techniques for data analysis.*(2nd ed.). Boca Raton : CRC Press.